

# Isoform specific gene auto-regulation via miRNAs: a case study on miR-128b and ARPP-21

Molly Megraw<sup>1,2,§</sup>, Praveen Sethupathy<sup>1,2,□</sup>, Kiranmai Gumireddy<sup>5</sup>, Shane T. Jensen<sup>3</sup>, Qihong Huang<sup>5,\*</sup>, and Artemis G. Hatzigeorgiou<sup>1,2,4,6,\*</sup>

<sup>1</sup>Center for Bioinformatics, <sup>2</sup> Department of Genetics, School of Medicine, <sup>3</sup>Department of Statistics, The Wharton School, <sup>4</sup> Department of Computer and Information Science, School of Engineering, University of Pennsylvania, <sup>5</sup> The Wistar Institute, Philadelphia, PA, USA and <sup>6</sup> Institute of Molecular Oncology, BSRC "Alexander Fleming", Athens, Greece.

§Present affiliation:

*Institute for Genome Sciences and Policy, Duke University, Durham, NC 27708 USA.*

□Present affiliation:

*Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892 USA.*

\*Contact: qhuang@wistar.org, hatzigeorgiou@fleming.gr

## ABSTRACT

In this study, we investigate whether miRNAs located within “host” protein-coding genes may regulate the expression of their host genes. We find that 43 of 174 miRNAs encoded within RefSeq genes are predicted to target their host genes. Statistical analysis of this phenomenon suggests that gene auto-regulation via miRNAs may be under positive selective pressure. Our analysis also indicates that several of the 43 miRNAs have a much lower expectation of targeting their host genes by chance than others. Among these examples, we identify miR-128b:ARPP-21 (cyclic AMP-regulated phosphoprotein, 21 kD) as a case in which both the miRNA and the target site are also evolutionarily conserved. We provide experimental support for this miRNA:target interaction via reporter silencing assays, and present evidence that this isoform specific gene auto-regulation has been preserved in vertebrate species in order to prevent detrimental consequences of ARPP-21 over-expression in brain.

## Introduction

MicroRNAs (miRNAs) are small non-coding RNAs, ~21 nucleotides (nt) long, which play a crucial role in gene regulatory networks. Hundreds of miRNAs have been characterized in eukaryotic organisms ranging from plants to humans [1]. The vast majority of these have been found recently with the advent of sensitive small RNA cloning techniques [2-5]. At the time of our computational analysis, there were 474 known human miRNA precursors. Animal miRNA precursors are most commonly located either outside of protein-coding genes (“intergenic”) or within the introns of protein-coding genes, and more rarely in UTR regions, coding exons, or exons of non-coding transcripts. miRNAs have been shown to guide the RNA Induced Silencing Complex (RISC) of proteins to specific target sites predominantly within the 3’ UTR of mRNAs in order to induce immediate cleavage, localization to P-bodies, or translational repression [6]. Evidence suggests that miRNAs are predicted to regulate up to one third of all protein-coding genes in humans [7].

In many cases, the translation of a protein-coding gene can elicit a self-limiting process in which the resulting protein inhibits the transcription of the same gene through interaction with transcription factors [8-10]. However, it has not yet been demonstrated whether a gene’s transcription process can directly inhibit its own translation process, or whether such auto-regulation can be isoform specific. Encouragingly, there has been a study which lends computational support to the idea that miRNA-mediated negative auto-regulatory feedback loops may exist in humans, and suggests that these loops belong to a class of circuits which are prevalent in mammalian gene network architecture [11]. Since both computational and experimental evidence is accumulating in support of the idea that intronic miRNAs are in fact spliced out and expressed along with their host genes [12-15], we investigated whether such miRNAs may regulate the expression of the host genes in which they are located. We then examined whether these potential auto-regulatory cases are subject to selective pressure, and provide the first experimental evidence for an isoform-specific case which may be linked to neurodegenerative disease.

## Results

### Detecting Gene Auto-regulation via miRNAs

Using miRNA target prediction, we identified 43 miRNAs which are predicted to target their own host genes. For comparison, we also simulated the number of miRNAs which would target their own host genes if these miRNAs were randomly distributed among 3’ UTRs of protein-coding genes (see Materials and Methods). The observed value of 43 potential auto-regulatory cases lies at the 95th percentile of the randomization distribution (Supplementary Figure 1), suggesting that the phenomenon of host-gene targeting by miRNAs may be under positive selective pressure.

We also calculated the chance that each individual miRNA would target its own host gene. We find that a number of predicted auto-regulatory miRNAs have a comparatively low miRNA-specific chance of randomly targeting their own 3’ UTRs. We see from Table 1 that among predicted auto-regulatory cases involving a conserved target site, hsa-miR-128b is estimated to have only ~15% chance of randomly targeting the 3’ UTR of its host gene *ARPP-21*, whereas the probabilities of miR-661 or miR-488 randomly targeting their hosts are 31% and 51% respectively.

Additionally, both miR-128b and its *ARPP-21* target site are conserved through many species, suggesting that auto-regulation may provide an evolutionarily old and beneficial mechanism in this case. For these reasons, we chose the hsa-miR-128b:*ARPP-21* interaction for subsequent experimental testing.

**Table 1.** Predicted auto-regulatory cases with conserved target site

miRNA	Refseq id	Gene symbol	P_random*
miR-128b	NM_016300 NM_201380, NM_201379, NM_201378,	<i>ARPP-21</i>	0.148
miR-661	NM_000445	<i>PLEC1</i>	0.312
miR-488	NM_004319	<i>ASTN1</i>	0.515

Cases in which a miRNA located within a RefSeq gene is predicted to target the 3' UTR of its host gene. Cases with a conserved target site are shown here (all are intronic), please see Supplementary Table 1 for a complete display of all 43 predicted auto-regulatory cases.

\*P\_random is the probability that a miRNA targets its own host gene at random, computed as described in the Materials and Methods section.

## miR-128b Targets the Long Isoform of Host gene *ARPP-21*

To investigate this prediction in the laboratory, we used the reporter silencing assay to determine whether hsa-miR-128b is able to repress *ARPP-21* protein production. This assay is a standard in vitro experimental technique for determining whether predicted miRNA:target binding interactions are capable of reducing gene expression [16, 17]. In this experiment, two different versions of the *ARPP-21* 3'UTR- one with the hsa-miR-128b target site and one with this site deleted- were inserted downstream of luciferase “reporter” genes. Both of these constructs were transfected into cells that also contained hsa-miR-128b. The idea behind this assay is that if the amount of protein produced by the luciferase reporter construct containing the miR-128b target site were significantly reduced in the presence of miR-128b (as compared with the amount of protein produced when the target site is deleted), this would provide evidence that miR-128b can induce repression of *ARPP-21* under the conditions of the experiment. Figure 2 shows the results of the assay, indicating that miR-128b can in fact repress its host gene, the longer of two isoforms of *ARPP-21*. Supplementary Figure 2 illustrates the genomic configuration of miR-128b, the two isoforms of *ARPP-21*, and the 3'UTR of the longer isoform targeted by miR-128b.

## Discussion

In the experimentally supported case of miR-128b:*ARPP-21*, we observed that miR-128b resides exclusively in the longer of two *ARPP-21* isoforms. It is not extremely rare for a miRNA to be present within an intron of one isoform and not another. In fact, this is the case for 26 out of 104 known human intronic miRNAs which occur in genes that have multiple isoforms (Supplementary Table 2). However, miR-128b:*ARPP-21* is a

particularly interesting case because although miR-128b and *ARPP-21* are highly expressed in human brain [18-20], only the longer isoform contains miR-128b target sites. This suggests that miR-128b targeting of *ARPP-21* may have evolved to prevent the accumulation of only the longer isoform in human brain.

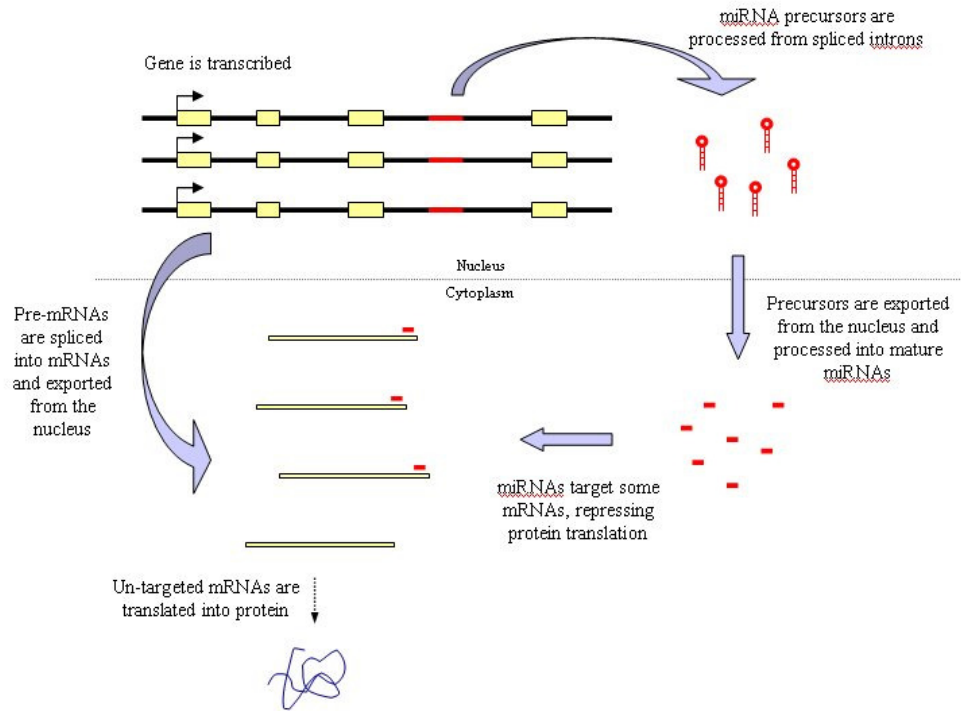
Although *ARPP-21* has not been extensively studied, there is evidence that it is highly expressed in dopamine-innervated brain regions [21-23]. Provocatively, a recent study using gene expression data to show that miRNA-mediated feedback and feedforward circuits are recurrent in human and mouse also concludes that brain-enriched miRNAs tend to target brain-enriched genes in these organisms [11]. The study suggests that these miRNAs could be involved in neuronal homeostasis. Furthermore, a recent patent suggests that the long isoform of *ARPP-21* is significantly differentially expressed in the brains of Alzheimer's patients as compared with those of age-matched healthy individuals [24]. One possible cause of this differential regulation may be the abrogation of the miR-128b:*ARPP-21* regulatory mechanism.

Of course it should be noted that miR-128b is not the only miRNA which may target the long isoform of *ARPP-21*. miR-128a, miR-107, miR-103, miR-9, and miR-29a are all miRNAs which are predicted to target *ARPP-21* and reported to be expressed in human brain by microarray studies, though only miR-128 and miR-9 are highly expressed [18, 19, 25-27]. This suggests that perhaps either miR-128b acts in concert with other miRNAs to substantially repress *ARPP-21* levels, or perhaps only miR-128b expression is highly specific to the brain regions where fine-tuned control of *ARPP-21* expression is required. In either case, the maintenance of this auto-regulatory system strongly suggests a mechanism which benefits vertebrate organisms by balancing the expression of these isoforms of *ARPP-21*.

## Methods

### Identifying Cases of Host-Gene Targeting by miRNAs

We used the database miRGen [28] to identify all human miRNAs encoded within a RefSeq host protein-coding gene. We identified exactly 174 such miRNAs, derived from 163 precursors (some precursors yield two miRNAs). To determine whether any of these miRNAs target their hosts, we applied a target prediction program [29] that predicts all currently known categories of miRNA target sites, including both 5'-dominant (according to rules from [7]) and 3'-compensatory target interactions (according to rules from [30]). Using this program, we predicted that 43 of the 174 miRNAs target their host genes (Supplementary Table 1). 35 of these are located within introns (Figure 1), three are located within a 5' UTR, one is located within a 3' UTR, and one overlaps an exon.



**Fig. 1.** An illustration of gene auto-regulation via an intronic miRNA: (A) Gene is transcribed, miRNA precursors are processed from spliced introns. (B) Pre-mRNAs are spliced into mRNAs and exported from the nucleus. (C) miRNA precursors are exported from the nucleus and processed into mature miRNAs. (D) miRNAs target some mRNAs, repressing protein translation. (E) Un-targeted mRNAs are translated into protein.

To determine how many of the 43 potential human auto-regulatory cases are also present in other species, we used the cross-species target conservation filter implemented in DIANA-microT 2.0. We found that the target sites in only three cases are conserved between human, chimp, mouse, rat, and dog – hsa-miR-128b:*ARPP-21*, hsa-miR-661:*PLEC1*, and hsa-miR-488:*ASTN1* (Supplementary Table 1). Among these cases, only hsa-miR-128b:*ARPP-21* was also predicted by other publicly available target prediction programs: PicTar [31] and TargetScanS [7]. The hsa-miR-488:*ASTN1* and the hsa-miR-661:*PLEC1* interactions were not included in the predictions of other programs because hsa-miR-488 and hsa-miR-661 were not an experimentally verified human miRNAs at the time of their publication and therefore were not incorporated into their genome-wide target searches.

### Null distribution of the Auto-regulatory miRNAs

How many auto-regulatory miRNAs would we observe under the assumption of random targeting? We first computed the “expected number of target sites per base pair” for each miRNA. This is the ratio of the number of predicted target sites in 3’ UTRs over the total length of all 3’ UTRs for that miRNA in the human genome.

The miRNA-specific expected number of target sites per base pair was then used to repeatedly simulate a random map of target genes for each miRNA. Each map was constructed by randomly adding miRNA target sites across gene 3’ UTRs. The

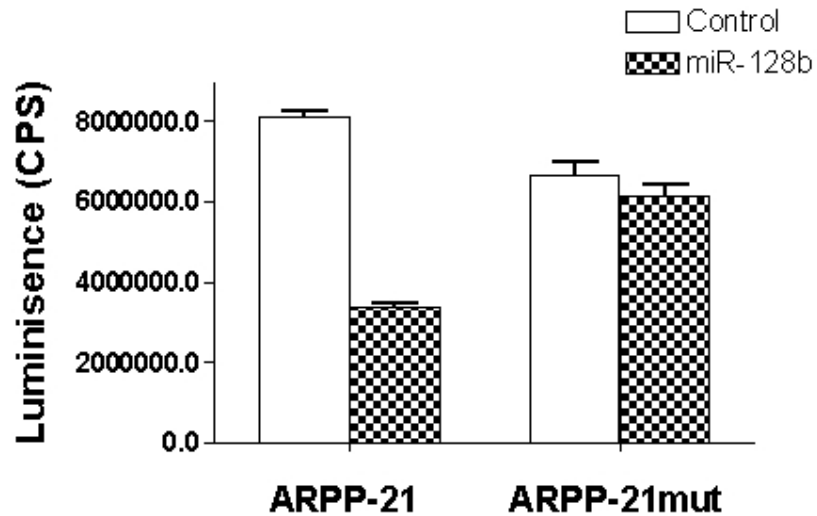
probability of a particular miRNA target site being added was equal to the “expected number of target sites per base pair” for that miRNA. For each of these simulated maps of target sites, we tabulated the number of auto-regulatory cases where a miRNA targeted its own 3' UTR. We repeated this simulation 10000 times, which gave us the distribution of the number of auto-regulatory miRNAs under the null hypothesis of random allocation. This null distribution is given in Supplementary Figure 1.

When compared to this randomization distribution, we found that the 43 observed potential auto-regulatory cases is quite high (95<sup>th</sup> percentile) when compared to range of values that we calculated in our simulation experiment. This may suggest that the phenomenon of auto-regulation is undergoing some degree of global positive selection in humans. It is interesting to note that those miRNAs which are least evolutionarily conserved were observed to account for a larger portion of predicted auto-regulatory cases than highly conserved miRNAs, perhaps indicating that the mechanism of auto-regulation confers species-specific benefits in higher organisms. This observation is explored in further detail in the Supplementary Methods.

In addition to this global simulation we also calculated the null probability for each miRNA. Under the null hypothesis of random allocation, the target site for a particular miRNA will appear at a specific location in the 3' UTR of its host gene with probability equal to the “expected number of target sites per base pair” for this particular miRNA. For each miRNA separately, the binomial distribution was used to calculate the probability of observing one or more target sites in the host gene 3' UTR, with the underlying binomial probability equal to the expected number of target sites per base pair and the number of trials equal to the number of base pairs in that 3' UTR. The underlying assumptions of the Binomial distribution (independence and equal probabilities between trials) are justified by our null hypothesis of completely random allocation. These miRNA specific null probabilities are given in the Table 1 column labeled “P\_random”.

### **Experimental Evidence that miR-128b Targets ARPP-21**

To investigate the prediction that miR-128b is capable of targeting *ARPP-21* in vitro, we inserted two different *ARPP-21* 3' UTRs downstream of luciferase reporter genes to create two distinct reporter constructs: (1) full length long isoform *ARPP-21* 3' UTR with the hsa-miR-128b target site, and (2) full length *ARPP-21* 3' UTR with the hsa-miR-128b target site deleted. We infected HEK293 cells with viruses containing these constructs (see Supplementary Methods). In the presence of hsa-miR-128b, the expression of the luciferase containing the wild-type *ARPP-21* 3' UTR was significantly reduced (Figure 2) confirming the function of hsa-miR-128b on the *ARPP-21* target gene. In contrast, the expression of the  $\Delta$ hsa-miR-128b target site luciferase construct remained unchanged in presence or absence of hsa-miR-128b (Figure 2), implying that the target site is directly responsible for the repression of luciferase containing wild-type *ARPP-21* 3' UTR in the presence of hsa-miR-128b.



**Fig. 2.** Luciferase activity of HEK293 cells containing the combination of human miR-128 or control vector and luciferase reporter of the 3' UTR of ARPP-21 or ARPP-21 mutant with miR-128 target site deleted. In the chart, the luminescence unit is Counts Per Second which measures the luciferase signal. Three independent experiments were performed with triplicates; error bars represent standard deviation.

#### **Acknowledgements**

A.G.H., M.M., and P.S. were supported by an NSF Career Award (DBI-0238295). P.S. was also supported by a predoctoral NIH training grant (5T32GM008216). S.T.J. was supported by a University of Pennsylvania Research Foundation Grant. Q.H. and K.G. were supported by the Commonwealth Universal Research Enhancement Program, Pennsylvania Department of Health, and Wistar startup fund.

## References

- 1 Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A. and Enright, A. J. (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* **34**, D140-144
- 2 [Lagos-Quintana, M., Rauhut, R., Lendeckel, W. and Tuschl, T. \(2001\) Identification of novel genes coding for small expressed RNAs. \*Science.\* \*\*294\*\*, 853-858](#)
- 3 Lau, N. C., Lim, L. P., Weinstein, E. G. and Bartel, D. P. (2001) An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science.* **294**, 858-862
- 4 [Lee, R. C. and Ambros, V. \(2001\) An extensive class of small RNAs in \*Caenorhabditis elegans\*. \*Science.\* \*\*294\*\*, 862-864](#)
- 5 Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M. and Dreyfuss, G. (2002) miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev.* **16**, 720-728
- 6 [Kloosterman, W. P. and Plasterk, R. H. \(2006\) The diverse functions of microRNAs in animal development and disease. \*Dev Cell.\* \*\*11\*\*, 441-450](#)
- 7 [Lewis, B. P., Burge, C. B. and Bartel, D. P. \(2005\) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. \*Cell.\* \*\*120\*\*, 15-20](#)
- 8 Paukku, K. and Silvennoinen, O. (2004) STATs as critical mediators of signal transduction and transcription: lessons learned from STAT5. *Cytokine Growth Factor Rev.* **15**, 435-455
- 9 [Haberland, M., Arnold, M. A., McAnally, J., Phan, D., Kim, Y. and Olson, E. N. \(2007\) Regulation of HDAC9 gene expression by MEF2 establishes a negative-feedback loop in the transcriptional circuitry of muscle differentiation. \*Mol Cell Biol.\* \*\*27\*\*, 518-525](#)
- 10 [Scheffe, J. H., Menk, M., Reinemund, J., Effertz, K., Hobbs, R. M., Pandolfi, P. P., Ruiz, P., Unger, T. and Funke-Kaiser, H. \(2006\) A novel signal transduction cascade involving direct physical interaction of the renin/prorenin receptor with the transcription factor promyelocytic zinc finger protein. \*Circ Res.\* \*\*99\*\*, 1355-1366](#)
- 11 [Tsang, J., Zhu, J. and van Oudenaarden, A. \(2007\) MicroRNA-mediated feedback and feedforward loops are recurrent network motifs in mammals. \*Mol Cell.\* \*\*26\*\*, 753-767](#)
- 12 [Baskerville, S. and Bartel, D. P. \(2005\) Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. \*Rna.\* \*\*11\*\*, 241-247](#)
- 13 [Du, G., Yonekubo, J., Zeng, Y., Oisami, M. and Frohman, M. A. \(2006\) Design of expression vectors for RNA interference based on miRNAs and RNA splicing. \*FEBS J.\* \*\*273\*\*, 5421-5427](#)
- 14 [Lin, S. L., Miller, J. D. and Ying, S. Y. \(2006\) Intronic MicroRNA \(miRNA\). \*J Biomed Biotechnol.\* \*\*2006\*\*, 26818](#)
- 15 [Lin, S. L. and Ying, S. Y. \(2006\) Gene silencing in vitro and in vivo using intronic microRNAs. \*Methods Mol Biol.\* \*\*342\*\*, 295-312](#)
- 16 [Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P. and Burge, C. B. \(2003\) Prediction of mammalian microRNA targets. \*Cell.\* \*\*115\*\*, 787-798](#)



- 17 Kiriakidou, M., Nelson, P. T., Kouranov, A., Fitziev, P., Bouyioukos, C., Mourelatos, Z. and Hatzigeorgiou, A. (2004) A combined computational-experimental approach predicts human microRNA targets. *Genes Dev.* **18**, 1165-1178
- 18 Barad, O., Meiri, E., Avniel, A., Aharonov, R., Barzilai, A., Bentwich, I., Einav, U., Gilad, S., Hurban, P., Karov, Y., Lobenhofer, E. K., Sharon, E., Shibolet, Y. M., Shtutman, M., Bentwich, Z. and Einat, P. (2004) MicroRNA expression detected by oligonucleotide microarrays: system establishment and expression profiling in human tissues. *Genome Res.* **14**, 2486-2494
- 19 Liu, C. G., Calin, G. A., Meloon, B., Gamliel, N., Sevignani, C., Ferracin, M., Dumitru, C. D., Shimizu, M., Zupo, S., Dono, M., Alder, H., Bullrich, F., Negrini, M. and Croce, C. M. (2004) An oligonucleotide microchip for genome-wide microRNA profiling in human and mouse tissues. *Proc Natl Acad Sci U S A.* **101**, 9740-9744
- 20 Su, A. I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K. A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., Cooke, M. P., Walker, J. R. and Hogenesch, J. B. (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci U S A.* **101**, 6062-6067
- 21 Brene, S., Lindfors, N., Ehrlich, M., Taubes, T., Horiuchi, A., Kopp, J., Hall, H., Sedvall, G., Greengard, P. and Persson, H. (1994) Expression of mRNAs encoding ARPP-16/19, ARPP-21, and DARPP-32 in human brain tissue. *J Neurosci.* **14**, 985-998
- 22 Ivkovic, S. and Ehrlich, M. E. (1999) Expression of the striatal DARPP-32/ARPP-21 phenotype in GABAergic neurons requires neurotrophins in vivo and in vitro. *J Neurosci.* **19**, 5409-5419
- 23 Caporaso, G. L., Bibb, J. A., Snyder, G. L., Valle, C., Rakhilin, S., Fienberg, A. A., Hemmings, H. C., Nairn, A. C. and Greengard, P. (2000) Drugs of abuse modulate the phosphorylation of ARPP-21, a cyclic AMP-regulated phosphoprotein enriched in the basal ganglia. *Neuropharmacology.* **39**, 1637-1644
- 24 Hipfel, R., Hanes, J., Von Der Kammer, H. and Pohlner, J. (2006) Camp-regulated phosphoprotein for diagnostic and therapeutic use in neurodegenerative diseases. Patent No. 20060024305
- 25 Bartel, D. P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell.* **116**, 281-297
- 26 Sempere, L. F., Freemantle, S., Pitha-Rowe, I., Moss, E., Dmitrovsky, E. and Ambros, V. (2004) Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. *Genome Biol.* **5**, R13
- 27 Shingara, J., Keiger, K., Shelton, J., Laosinchai-Wolf, W., Powers, P., Conrad, R., Brown, D. and Labourier, E. (2005) An optimized isolation and labeling platform for accurate microRNA expression profiling. *Rna.* **11**, 1461-1470
- 28 Megraw, M., Sethupathy, P., Corda, B. and Hatzigeorgiou, A. G. (2007) miRGen: a database for the study of animal microRNA genomic organization and function. *Nucleic Acids Res.* **35**, D149-155

- 29 Kawahara, Y., Zinshteyn, B., Sethupathy, P., Iizasa, H., Hatzigeorgiou, A. G. and Nishikura, K. (2007) Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. *Science*. **315**, 1137-1140
- 30 [Brennecke, J., Stark, A., Russell, R. B. and Cohen, S. M. \(2005\) Principles of microRNA-target recognition. \*PLoS Biol.\* \*\*3\*\*, e85](#)
- 31 [Krek, A., Grun, D., Poy, M. N., Wolf, R., Rosenberg, L., Epstein, E. J., MacMenamin, P., da Piedade, I., Gunsalus, K. C., Stoffel, M. and Rajewsky, N. \(2005\) Combinatorial microRNA target predictions. \*Nat Genet.\* \*\*37\*\*, 495-500](#)